



Communication 1

**Innovation in the analysis of data for the construction of
epidemiological and predictive models with BIG DATA.
Application to colorectal models**



University of
Almería

Juan Manuel García-Torrecillas. Hospital Torrecárdenas, Almería

Jesús de la Fuente Arias. University of Almería

INTRODUCTION

■ Importance:

- CRC involves specialties in mode multidisciplinary
- Synergy in different hits: screening, treatment, prevention
- Two hard dependent factors: Mortality and Readmissions

■ Epidemiology

- Third most frequent malignant tumour
- 15% of tumoural incidence in Spain
- 25,000 new episodes/year in Spain
- Second etiology of tumoural death: 13,000 deaths/year

- Spain has no developed predictive models in mortality
- No specific models on clinical-administrative data sets
- PM por mortality in CCR
 - POSSUM
 - P-POSSUM
 - Cr-POSSUM
 - CGBI
 - Physiological Score
 - Patological Score

MODELLING PREDICTIVE MODELS

- Logistic Regression (Usual)
- Multilevel Logistic Regression (Unusual)
- New Explorations
 - Neural Networks (Francis, 2015. Readmissions)
 - Propensity Score: Propension and Matching
 - Structural Equations (SEM)
 - Bayesian focus

AVAILABLE SCORES IN CRC CURRENTLY

Panel 1: Physiological and operative factors for calculation of POSSUM, P-POSSUM, and CR-POSSUM scores

Physiological score

Age
Cardiac signs
Respiratory signs*
Electrocardiography*
Systolic blood pressure (mm Hg)
Pulse
Haemoglobin (g/100 mL)
White cell count ($\times 10^{12}$)*
Urea (mmol/L)
Sodium (mmol/L)*
Potassium*
Glasgow coma score*

Operative score

Operative severity
Multiple procedures*
Total blood loss (mL)*
Peritoneal soiling
Malignant disease status
Mode of surgery

*Measurement not included in calculation of CR-POSSUM.

POSSUM AND P&Cr-POSSUM

Panel 2: Equations for calculation of risk of death by predictive score

POSSUM

$\text{Log}_e [R/(1-R)] = -7.04 + (0.13 \times \text{physiological score}) + (0.16 \times \text{operative score})$

P-POSSUM

$\text{Log}_e [R/(1-R)] = -9.065 + (0.16 \times \text{physiological score}) + (0.15 \times \text{operative score})$

CR-POSSUM

$\text{Log}_e [R/(1-R)] = -9.167 + (0.33 \times \text{physiological score}) + (0.30 \times \text{operative score})$

ACPGBI

$\text{Log}_e [R/(1-R)] = +(4.859 - \text{total score})$

R=predicted risk of death for POSSUM, P-POSSUM, and CR-POSSUM predictive scores, and for ACPGBI score.

Summary of model performance of the CR-POSSUM and P-POSSUM scoring systems according to ROC analysis

Score	AUC
CR-POSSUM	0.675
P-POSSUM	0.611

Comparison of model performance of P-POSSUM and CR-POSSUM scoring systems

Study	Model performance (validation set)		
	Discrimination ^a (%)	Calibration ^b	O/E mortality (%)
P-POSSUM	61.1	44.85, $P = 0.002$	1.62
CR-POSSUM	67.5	42.20, $P = 0.013$	1.50

O/E: observed to expected mortality.

^aDiscrimination is measured by the area under the receiver–operator characteristic curve (standard error). Higher values represent better model discrimination.

^bCalibration is measured by the Hosmer–Lemeshow χ^2 statistic (4 degrees of freedom). Smaller values represent better model calibration.

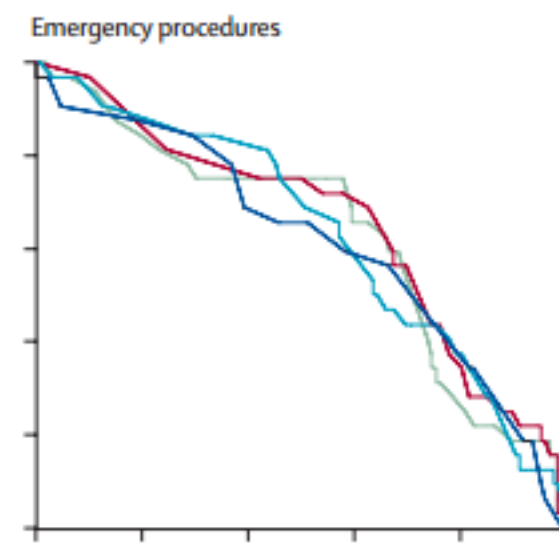
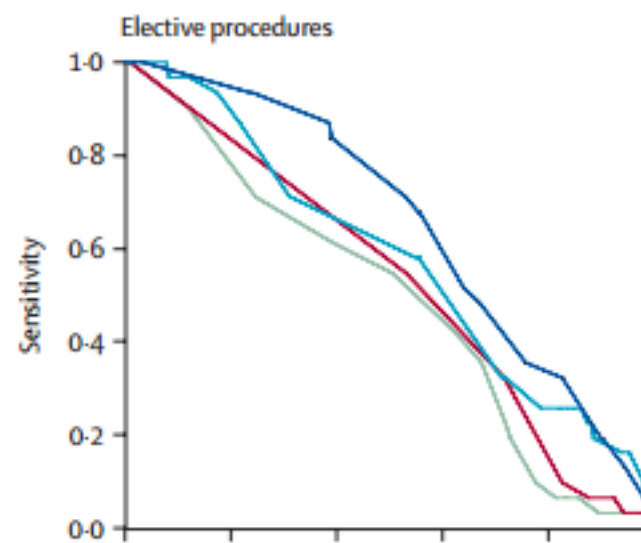
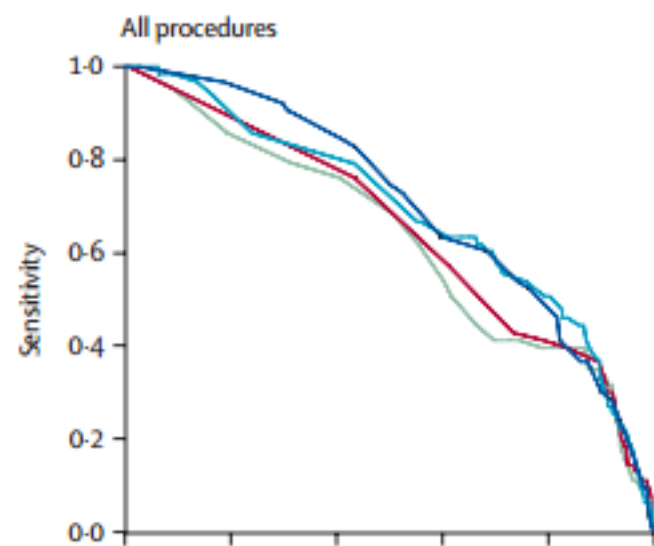
	Score
Age (years)	
<65	0
65-74	0.7
75-84	1.1
85-94	1.3
≥95	2.6
Cancer resected	
ASA I	0
ASA II	0.8
ASA III	1.6
ASA IV-V	2.5
Cancer not resected	
ASA I	1.7
ASA II	1.8
ASA III	2.1
ASA IV-V	2.4
Cancer stage	
Dukes' stage A	0
Dukes' stage B	0
Dukes' stage C	0.2
Dukes' stage D	0.6
Operative urgency	
Elective	0
Urgent	0.8
Emergency	1.1

Table 1: Factors for calculation of ACPGBI score

ACPGBI score	Predicted 30-day mortality
0-0	0.8
0.1-0.4	0.9-1.1
0.5-0.8	1.3-1.7
0.9-1.2	1.9-2.5
1.3-1.6	2.8-3.7
1.7-2.0	4.1-5.4
2.1-2.4	6.0-7.9
2.5-2.8	8.6-11.3
2.9-3.2	12.3-16.0
3.3-3.6	17.4-22.1
3.7-4.0	23.9-29.8
4.1-4.4	31.9-38.7
4.5-4.8	41.1-48.5
4.9-5.2	51.0-58.4
5.3-5.6	60.8-67.7
5.7-6.0	69.9-75.8
6.1-6.4	77.6-82.4
6.5-6.8	83.8-87.4

Table 2: Chart for conversion of ACPGBI score to predicted 30-day mortality⁷

POSSUM
P-POSSUM
CR-POSSUM
ACPGBI



	ROC AUC (95% CI)			
	POSSUM	P-POSSUM	CR-POSSUM	ACPGBI
All surgery	0.627 (0.587–0.664), p=0.017	0.648 (0.608–0.684), p=0.071	0.686 (0.647–0.723), p=0.528	0.701 (0.663–0.737)
Elective surgery	0.488 (0.443–0.536), p=0.001	0.527 (0.479–0.573), p=0.006	0.575 (0.527–0.620), p=0.078	0.651 (0.605–0.620)
Emergency surgery	0.627 (0.546–0.702), p=0.870	0.650 (0.572–0.726), p=0.609	0.632 (0.553–0.708), p=0.729	0.614 (0.531–0.688)
Colorectal surgeons	0.543 (0.499–0.588), p=0.001	0.577 (0.533–0.621), p=0.004	0.625 (0.581–0.667), p=0.073	0.680 (0.638–0.721)
Non-colorectal surgeons	0.657 (0.565–0.748), p=0.950	0.679 (0.584–0.764), p=0.716	0.678 (0.584–0.764), p=0.611	0.653 (0.556–0.739)

p values obtained from non-parametric test^a for comparison of POSSUM, P-POSSUM, and CR-POSSUM with ACPGBI score.

Table 5: ROC curve analysis

#EpiColorrecto

El presente estudio ha sido posible en parte gracias a las becas PI16/01931 del Instituto de Salud Carlos III procedentes de fondos del Ministerio de Sanidad, Servicios Sociales e Igualdad (MSSSI) y Fondo Europeo de Desarrollo Regional (FEDER).



"Una manera de hacer Europa"

Grupo de Investigación

Epidemiología del Cáncer Colorrectal

Scores From Research Team

Research article

Open Access

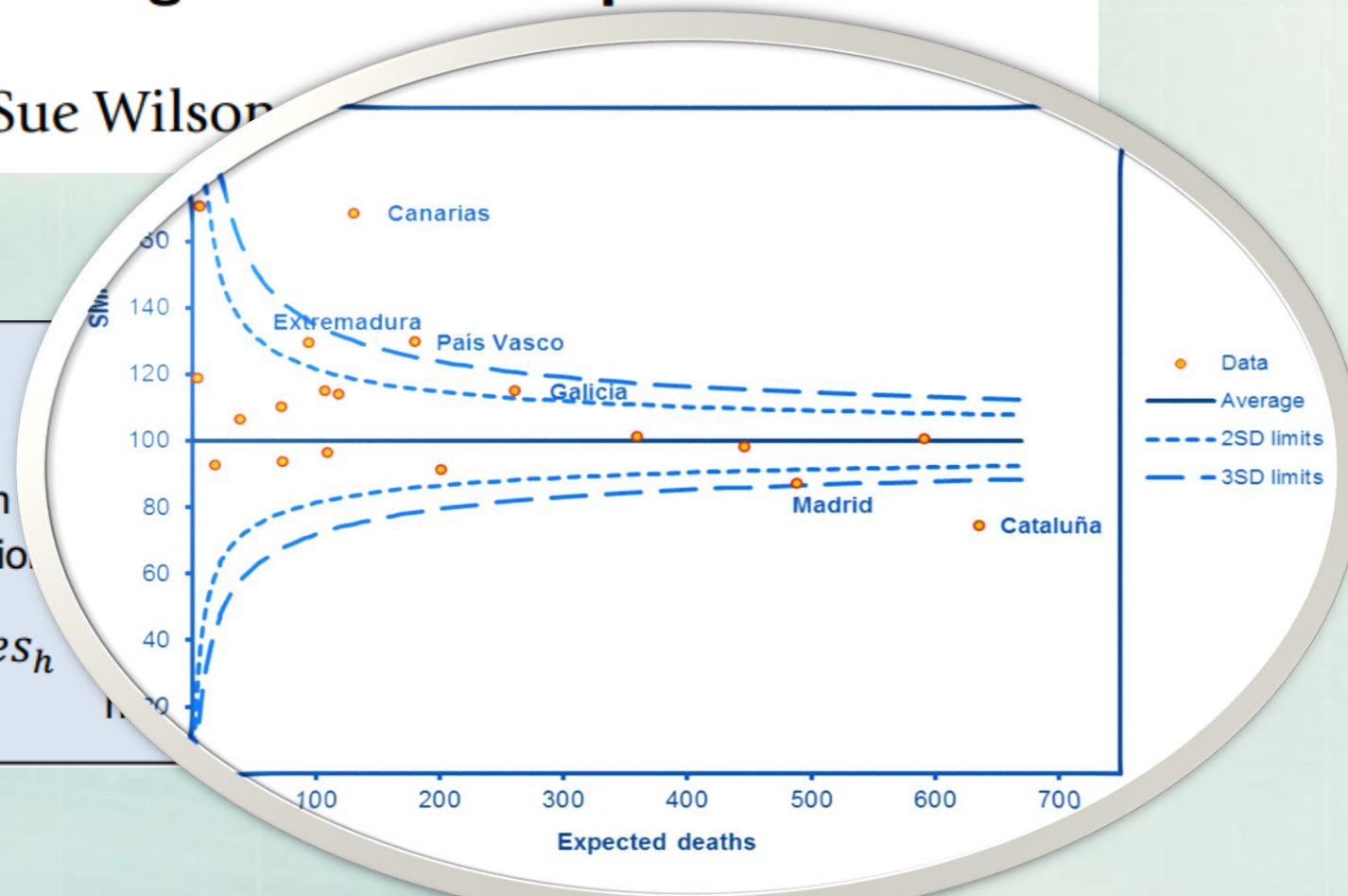
Standardisation of rates using logistic regression: a comparison with the direct method

Andrea K Roalfe*, Roger L Holder and Sue Wilson

$$HSMR_h = \frac{\text{Observed Deaths}_h}{\text{Predicted Deaths}_h} \quad \text{where,}$$

$\text{Observed Deaths}_h = \sum_j \text{Numerator}_h$ the sum of admissions

$\text{Predicted Deaths}_h = \sum_j \text{Predicted Probabilities}_h$

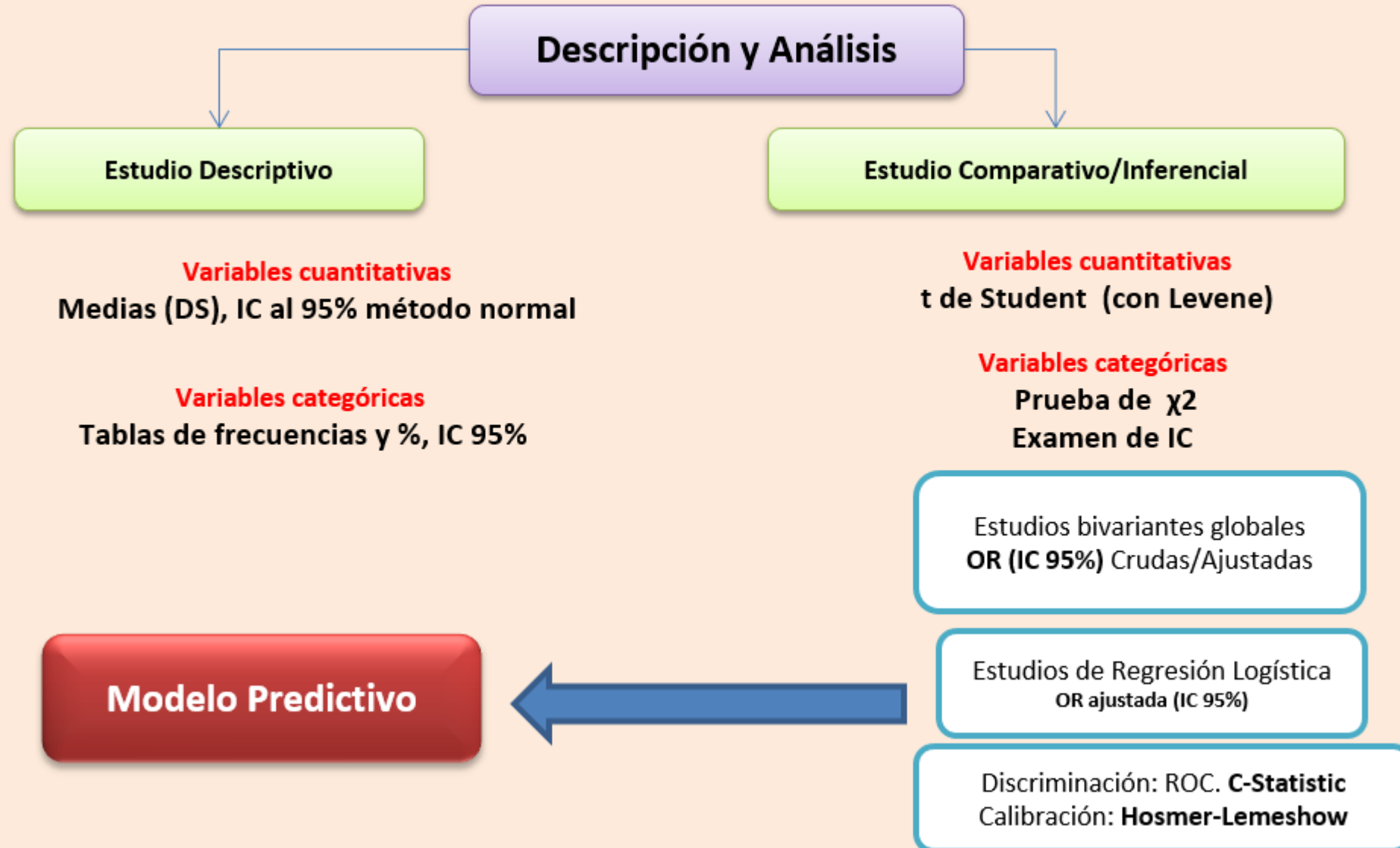


NON OUTLIERS PROLOGUED STAY

METHODS

- **Information source:** CMBD from MSCPS, 2008-2014 period
- **Design:** Descriptive, transversal, not concurrent
- **Codes :** CIE 153-154
- **Variables:** Age, Sex, Length of stay, Type of admission, DND, PND, comorbidities, readmission, PE (>20 days)
- **PE:** > percentile 90 (20 days)
- **Outlier:** $T2 = Q3 + 1,5(Q3 - Q1) = 29$ days

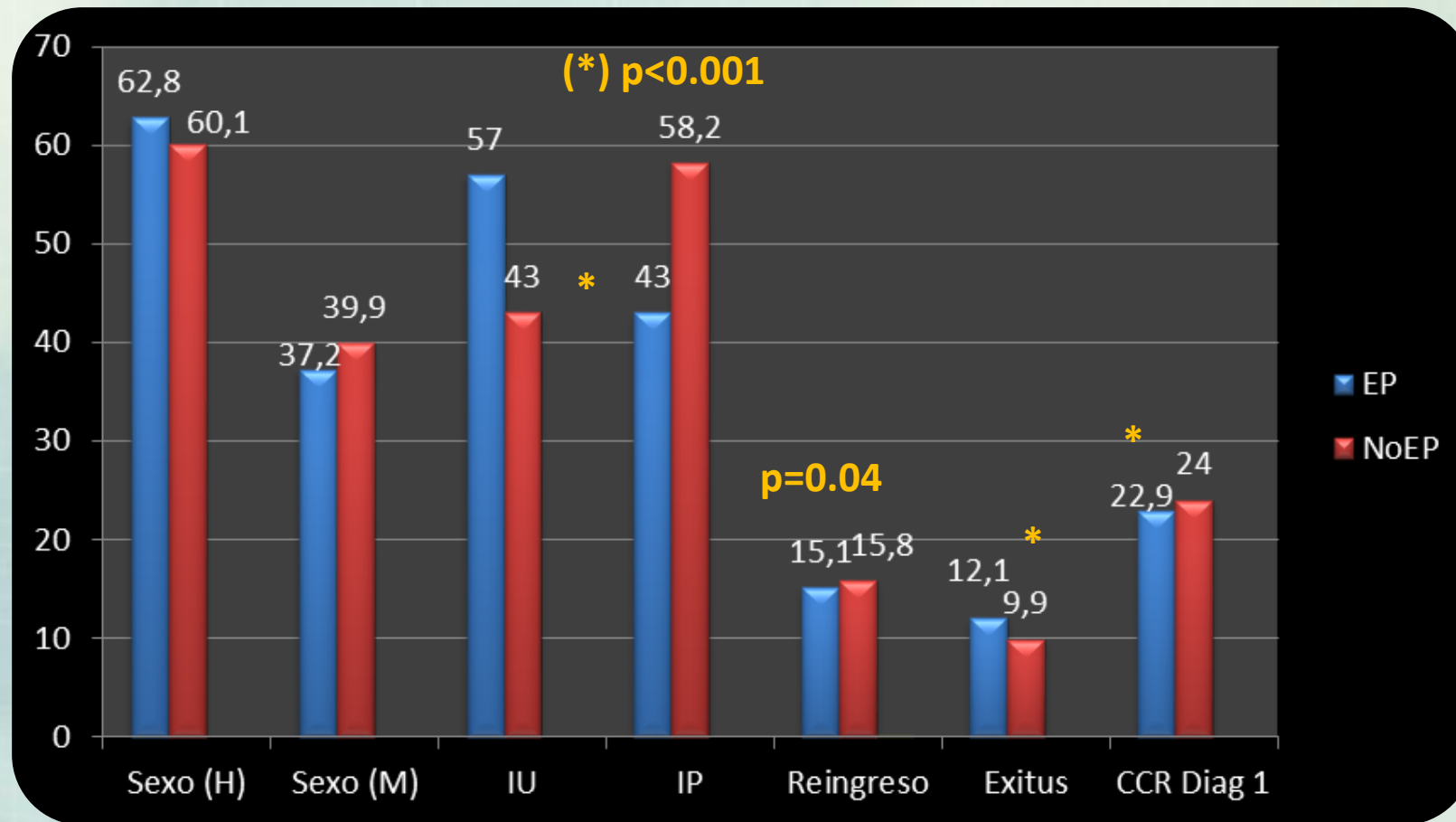
METHODS PRE-MODEL



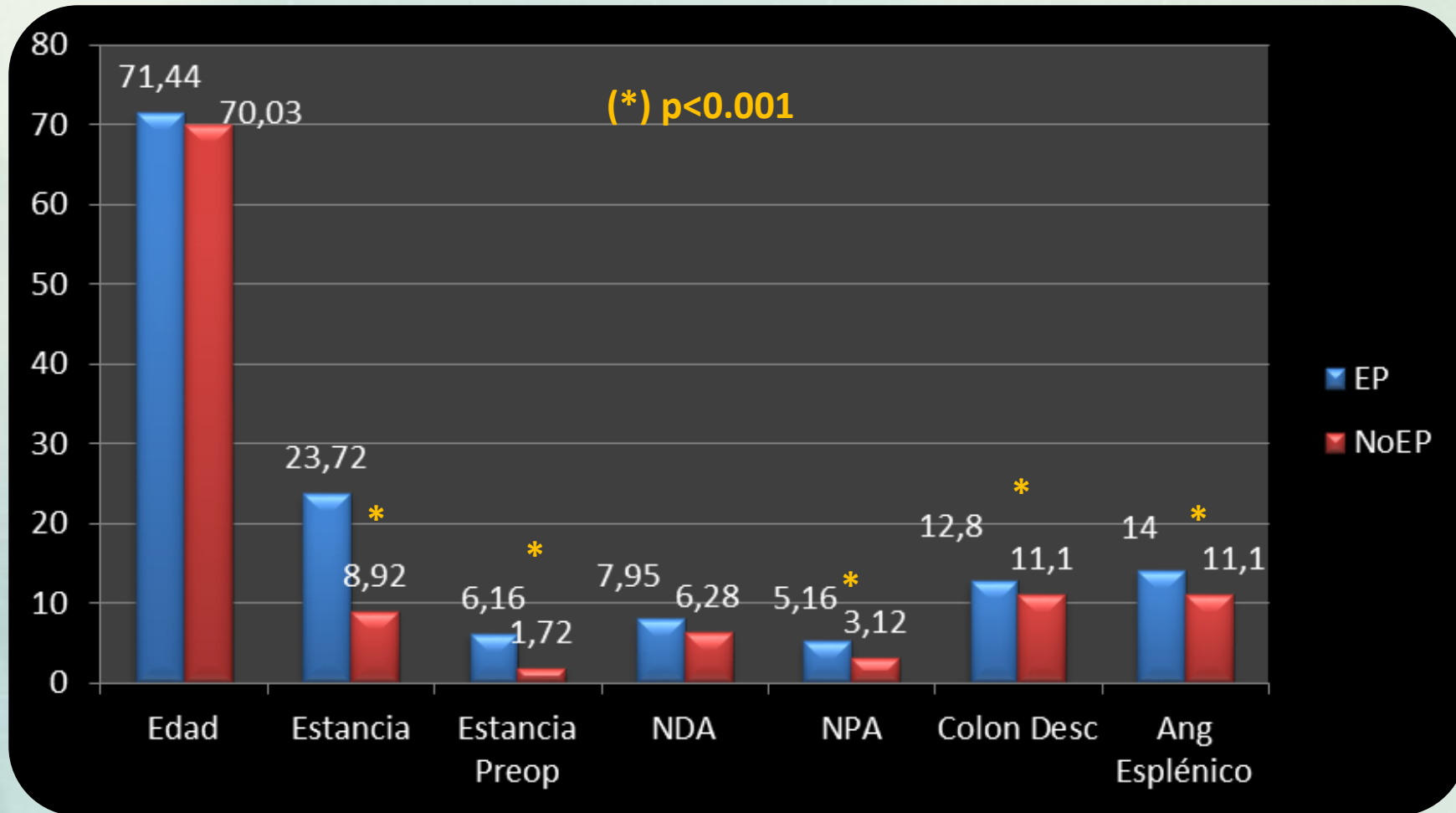
- 239.158 admissions, period 2008-14
- PE: 26.697

Estadísticos descriptivos						
	EP no Outlier	N	Mínimo	Máximo	Media	DS
No	Edad	212462	20	104	70,03	12,102
	Estancia	212462	0	19	8,92	4,480
	Estancia preoperatoria	100624	0	19	1,72	2,189
	Numero de diagnosticos	212462	1	14	6,28	3,312
	Numero de procedimientos	212462	0	20	3,12	2,528
	N válido (por lista)	100624				
Si	Edad	26697	20	100	71,44	11,716
	Estancia	26697	20	29	23,72	2,842
	Estancia preoperatoria	15909	0	29	6,16	6,659
	Numero de diagnosticos	26697	1	14	7,95	3,437
	Numero de procedimientos	26697	0	20	5,16	3,440
	N válido (por lista)	15909				

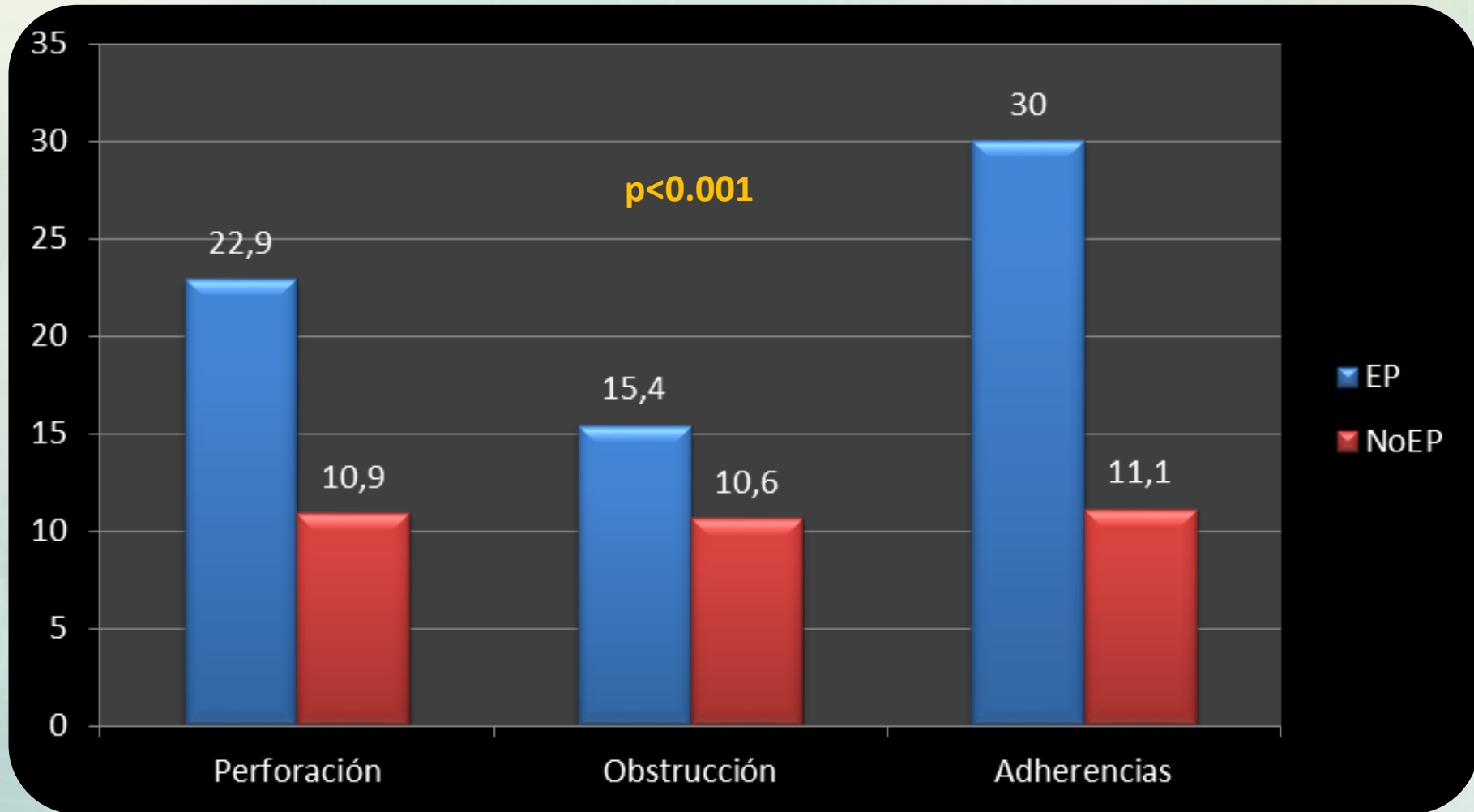
RESULTS I



RESULTS II



RESULTS III

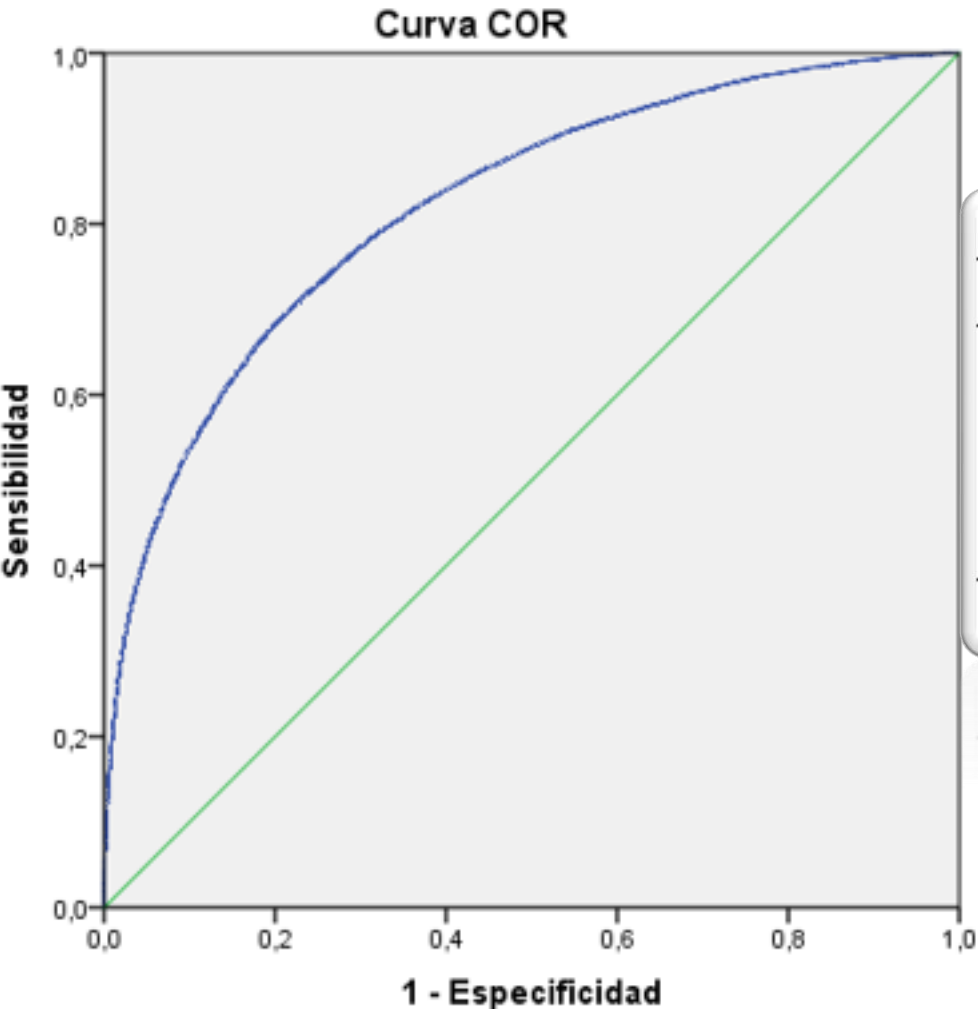


LOGISTIC MODEL FOR TRAINING COHORT

		B	EE	<u>Wald</u>	gl	Sig.	Odds Ratio	95% C.I. para OR	
								Inferior	Superior
	Edad	,007	,001	49,827	1	,000	1,007	1,005	1,009
	Número de diagnósticos	,114	,004	1033,255	1	,000	1,121	1,113	1,129
	Número de procedimientos	,130	,004	1368,825	1	,000	1,138	1,131	1,146
	Sexo(Hombre)	,224	,023	92,596	1	,000	1,251	1,195	1,309
	Ingreso Urgente	,237	,027	75,122	1	,000	1,267	1,201	1,337
	CCR_1 en posición primera	,216	,059	13,293	1	,000	1,241	1,105	1,394
	Situación Recto	,305	,026	134,802	1	,000	1,356	1,288	1,428
	Adherencias u obstrucción	,913	,106	74,422	1	,000	2,491	2,025	3,065
	Complicación de ostomía	,701	,096	53,768	1	,000	2,016	1,671	2,431
	Neumonía bacteriana	1,065	,084	162,190	1	,000	2,900	2,462	3,417
	Estancia preoperatoria	,229	,003	6262,787	1	0,000	1,257	1,250	1,265
	Constante	-4,931	,094	2751,114	1	0,000	,007		

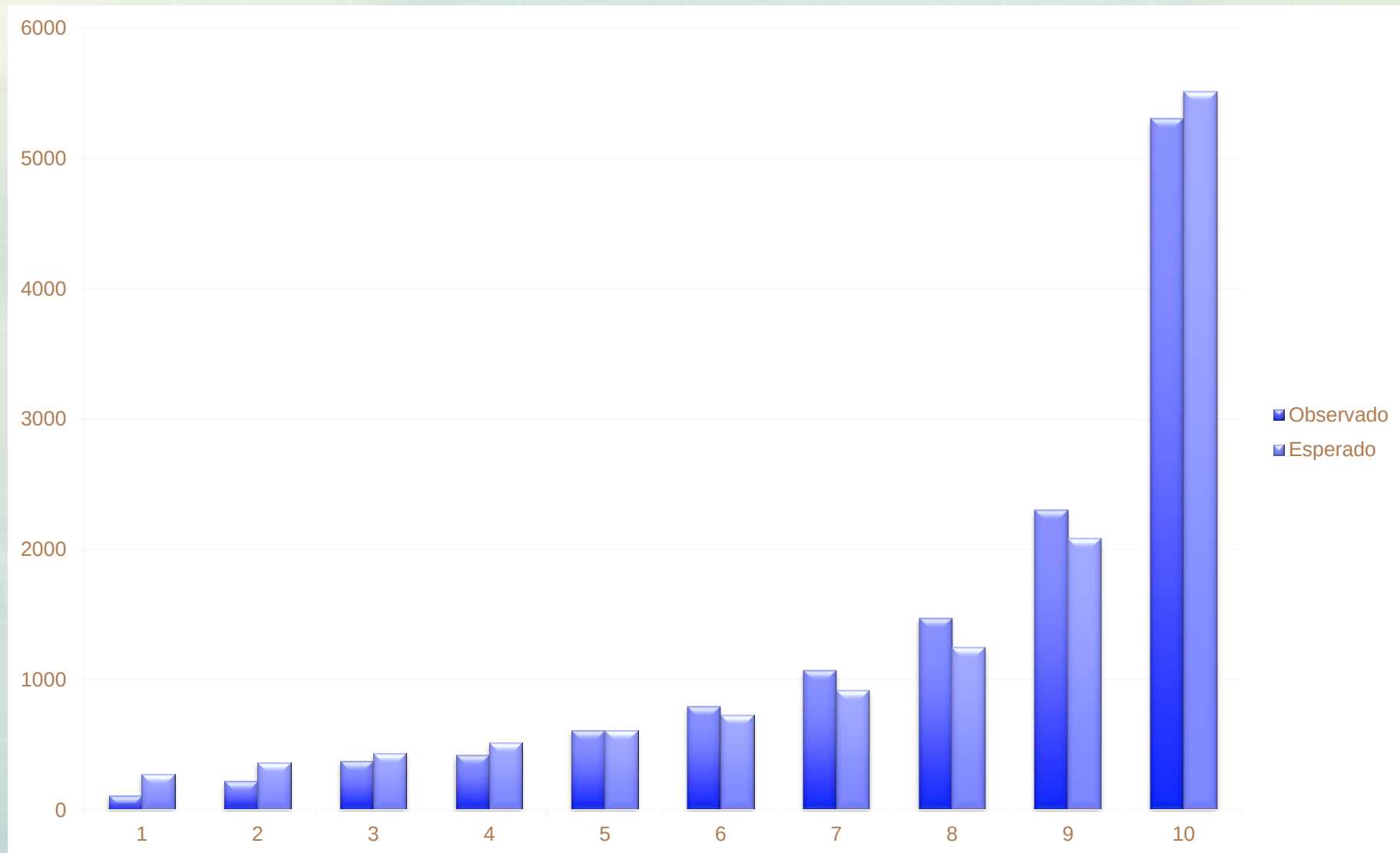
MODEL DISCRIMINATION

Área bajo la Curva. C Statistic (capacidad discriminativa del modelo)



Área bajo la curva				
Variables de resultado de prueba: Probabilidad pronosticada				
Área	Error estándar	Significación asintótica	95% de intervalo de confianza asintótico	
			Límite inferior	Límite superior
0,821	,002	,000	,817	,825

HOSMER LEMESHOW. CALIBRATION



$p < 0,005$

CONCLUSIONS

- Los pacientes con EP son una casuística especial dentro del grupo de hospitalizaciones por CCR
- Mayoritariamente son varones, que ingresan de modo urgente, requieren más esfuerzo procedimental (NPA), presentan más complicaciones específicas (perforación, obstrucción y especialmente adherencias) y mayor mortalidad.

Mortality. Extended Model

Variables in the study

Dependent Variable



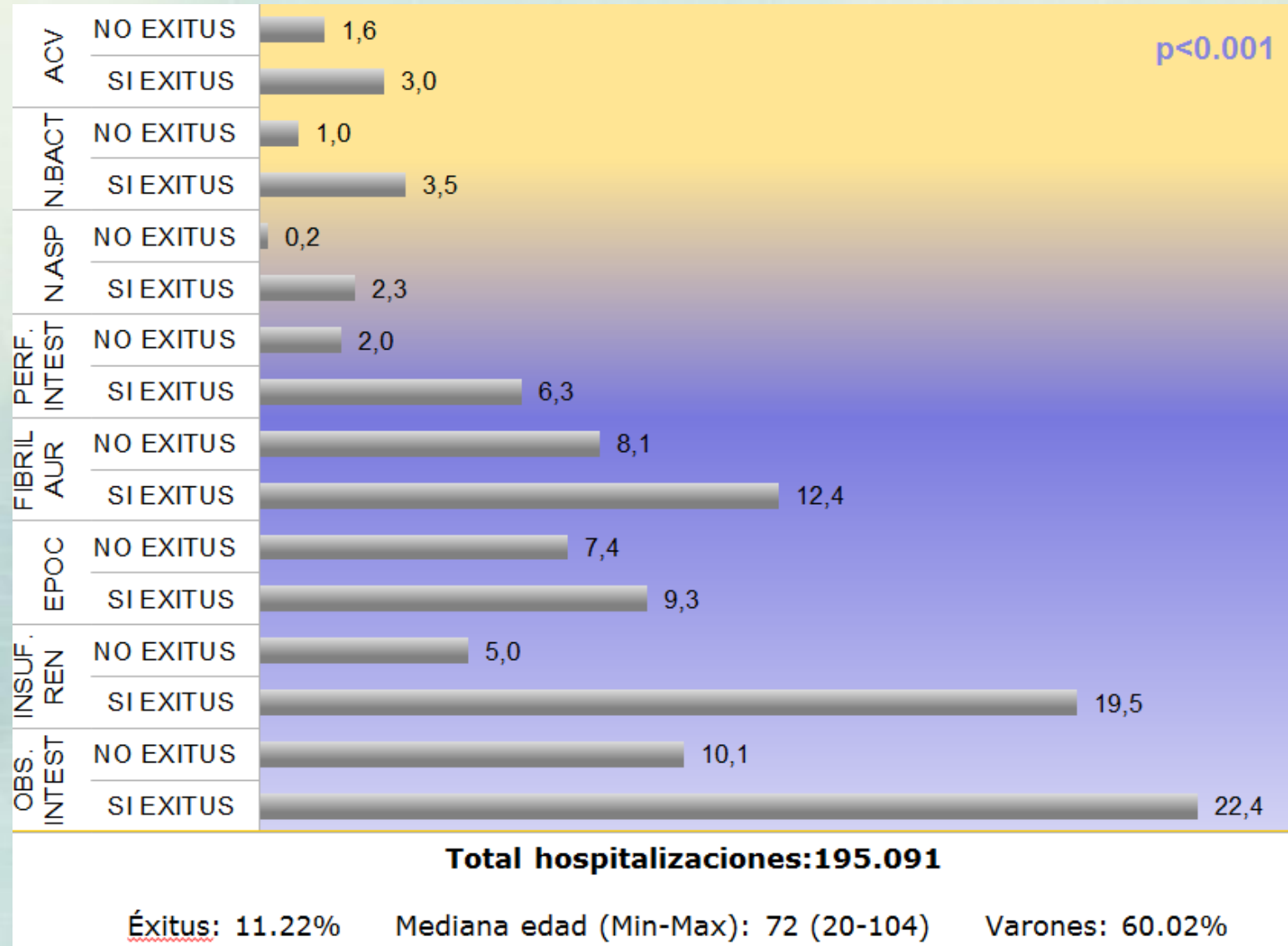
In Hospital Mortality

Independent Variables

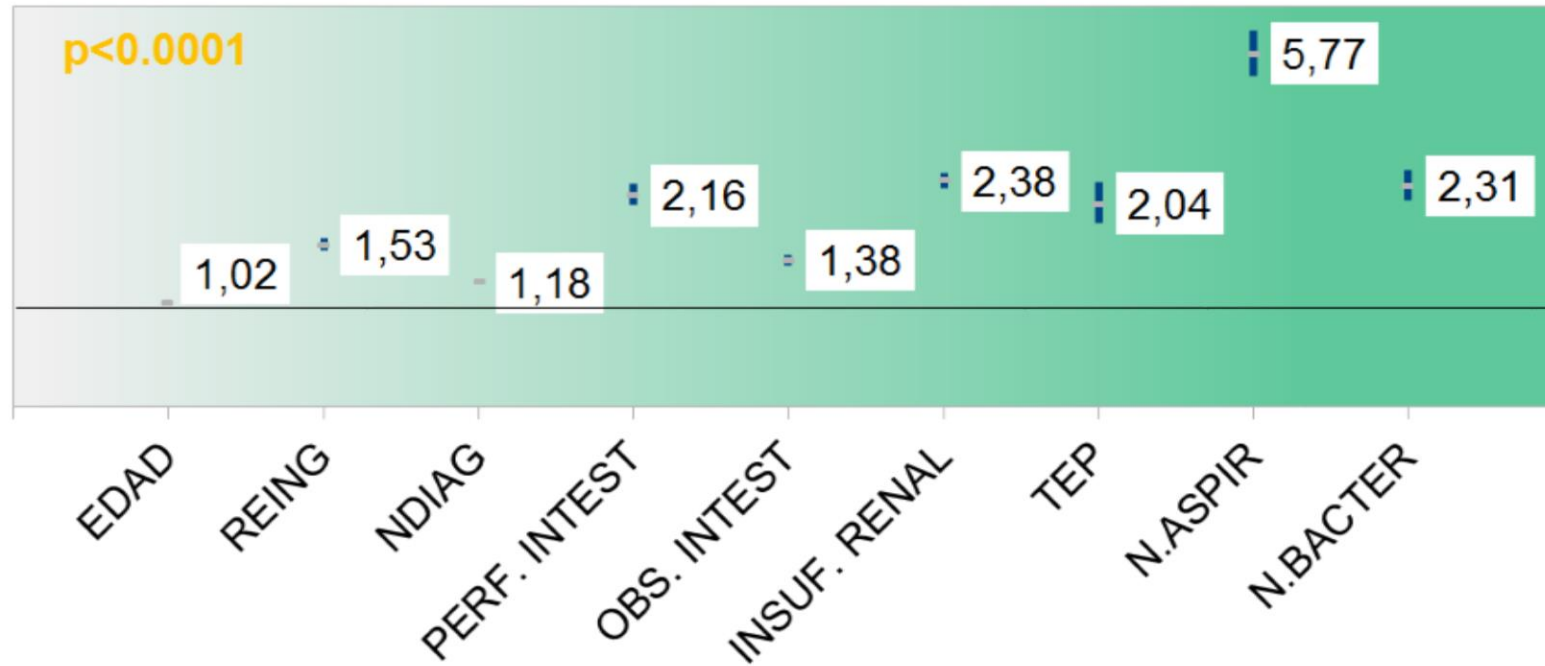
Age (years)
Year (2008-2014)
Sex (male-female)
AACC (Méth. Foster/SMR)
Surgery(Yes-No)
Tumour Localization
Financials

Dates: Born-Admission-Discharge
Type of Admission: (UI-PI)/ Type of discharge
CIE Diagnostic (14 places)
CIE Procedures (20 places)
Mortality
Clinical-Surgical Comorbidities
Clinical-Surgical Complications

Results



OR ajustadas para mortalidad por CCR



AUC=0.83 IC95% (0.827-0.833)

Adjusted by:

CCAA agruped by SMR
Procedure number

Cirugía
Type of AdmissionTipo

Length of Stay
Admission Dyslipidemia

PROSPECTIVE

- El uso de las grandes bases de datos clínico-administrativas, así como los datos en streaming junto a los nuevos métodos de análisis estadístico y el Big Data, debería revertir en modelos de atención calidad y gestión que optimicen el tratamiento de estos pacientes en el ámbito hospitalario y ambulatorio.
- La creación de MP con técnicas estadísticas avanzadas aportaría un plus de progresión en la metodología de abordaje del estudio epidemiológico, de manera específica en CRC pero, en definitiva, aplicable a muchas entidades